Predicting Road Quality Across Local Governments in Los Angeles County

CS231N: Deep Learning for Computer Vision

Andrea Vallebueno

Department of Statistics

Introduction

Background

- The maintenance of **quality roads** is essential to the social and economic inter-connectivity of human establishments. To study the dynamics of its provision by local governments, we require fine-grained measurements along a vast number of streets, a costly endeavor.
- We use street imagery to predict **pavement quality for block groups in Los Angeles County**. These inputs will be used in future research to study the provision of this public good.

Problem statement

- We leverage a bottom-up approach that generates **image-level predictions** of road quality and aggregates them to generate estimates at the **street** and **block group** levels.
- Our **image-level model** is a CNN that takes as input a 224 x 224 RGB image of a subset of a street's pavement, and outputs the predicted road quality class (serious, fair or good).

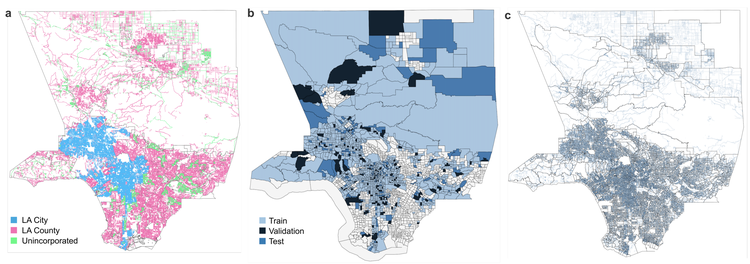


Figure 1. **a)** Network of +400,000 street segments in Los Angeles County. **b)** Census tracts containing labeled street segments are randomly allocated to training, validation and test splits. **c)** Block group-level road quality predictions are computed by sampling 10% of the street segments (blue) in each block group.

Dataset

We pair data on street-level road quality with street imagery to build a dataset of over 151,000 labeled images.

- **Road quality.** We collect data on the Pavement Condition Index (PCI), a continuous measure of the density of pavement distresses along a street, for streets from Los Angeles City's Public Works Department. The PCI is only available for streets in Los Angeles City and unincorporated areas of the county, and we uniformly bin it into **3 classes**.
- **Street imagery.** We query the Google Street View Static API for 16 images of the pavement at random locations along each street. Images are cropped and resized to 224x224.

Figure 2. Sample street segments from each class.

Dataset	Train split	Spatial Coverage	Segments	Images	Images per segment
Training	Train	1,135 census tracts	11,349	120,464	16
	Validation	141 census tracts	1,417	14,864	16
	Test	143 census tracts	1,692	17,712	16
Prediction	-	6,572 block groups	27,716	165,644	8

Table 1. Training and Prediction GSV datasets

Class imbalance in the dataset is quite severe due to the relative frequency of low and high-quality roads in Los Angeles County. The training set is comprised by Serious (7% of total), Fair (27%) and Good (66%) data points.

Methods

- We train an **image-level model** with a ResNet architecture that independently classifies the pavement quality of each image belonging to a street.
- We explore different model architectures to produce **segment-level predictions** (see Figure 3), and compute **block group predictions** by sampling 10% of the segments in a block group.

We experiment with the following methods to address class imbalance.

Sampling

- **Weighted sampling** by inverse class frequencies
- **Class-balanced datasets**, built by over-sampling low-frequency classes and under-sampling high-frequency classes
- **Two-stage training**, training on the imbalanced dataset and training the last layer on the balanced dataset

Loss functions

- **Class balanced loss** by Cui *et al.*, which weighs the loss according to the effective number of samples

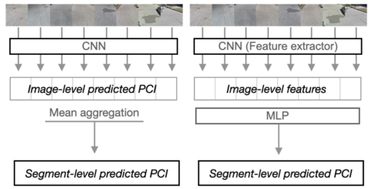


Figure 3. Explored model architectures to generate segment-level predictions.

Results & Analysis

Image-level performance

We evaluate model performance on the validation set using macro metrics, which equally weigh each class, including precision and recall.

Baseline model

- ResNet-18 leveraging pre-trained ImageNet weights, and trained with partial layer freezing, weighted sampling and cross-entropy loss

Best-performing (BP) model

- ResNet-50 trained end-to-end on the class-balanced dataset with cross-entropy loss
- Test-set accuracy of 54.7%, macro precision of 45.1% and macro recall of 51.7%

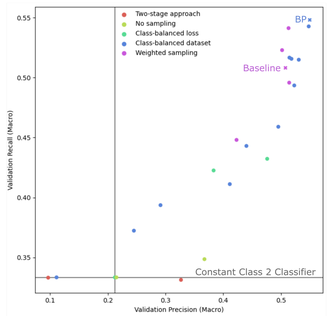


Figure 4. Best validation macro precision and recall.

We qualitatively analyze model performance using Class Activation Maps (Figure 5) and exploiting minor variations in images (Figure 6). Given the noise in the image-level labels, we perform error analysis and find cases of correct model predictions for mislabeled images (Figure 7).

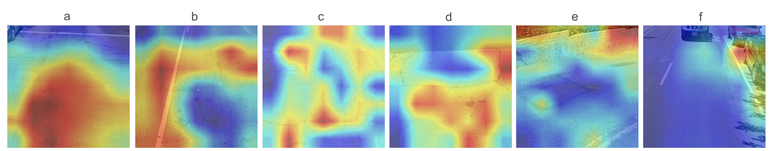


Figure 5. Class Activation Maps. **a, b)** Predictions for high road quality images driven by smooth activations across the distress-free pavement; **c, d)** Activations for lower quality roads appropriately driven by the presence of pavement cracking; **e, f)** Predictions remain sensitive to the presence of objects such as cars and shadowing.

Results & Analysis

Figure 6. Minor variations in camera angles and capture locations shed light on model prediction drivers such as shadow patterns and cracking positions.

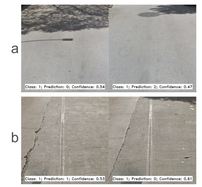
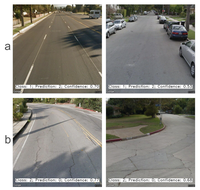


Figure 7. Noisy data points in the image-level dataset occasionally result in high confidence, correct predictions on mislabeled images.



Segment-level predictions

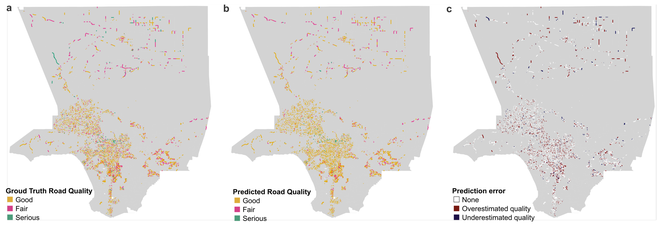


Figure 6. Segment-level road quality ground truth (**a**), model predictions (**b**) and prediction errors (**c**).

Block group-level predictions

The best-performing model achieves 78.8% accuracy, 39.1% macro precision and 38.7% macro recall on the test set, and spatially generalizes to block groups outside of Los Angeles City.

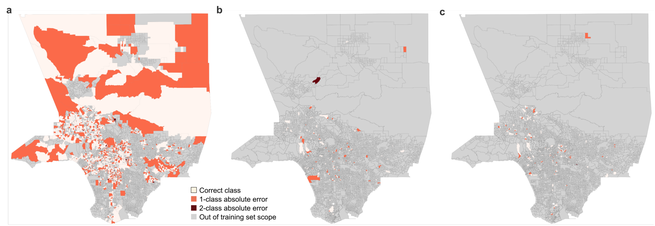


Figure 7. Block group-level predictions for block groups in the training (**a**), validation (**b**) and test (**c**) sets.

Conclusions and Future Work

- We implemented a bottom-up approach to generating block group-level predictions of road quality, an essential input to understanding the provision of this public good given its high measurement costs. An important limitation is the availability of street view imagery, which is collected at varying frequencies according to the location of interest.
- In future work we aim to address the **noise in the image-level road quality labels** and **severe class imbalance** through changes in model architecture to match the unit of labeling, re-labeling images using high confidence predictions, and a more efficient imagery collection process accounting for class imbalance at the segment sampling stage.